INFLUENCE OF SPECIFIC CONTACTS ON THE STABILITY AND STRUCTURE OF PROTEINS

Theory for the Perturbation of a Harmonic System

MEYER B. JACKSON
Department of Biology, University of California, Los Angeles, California 90024

ABSTRACT The question of how specific contacts within a protein influence its stability and structure is examined within a formal theoretical framework. A mathematical model is developed in which the potential energy of a protein is taken as a harmonic expansion of all of its internal or normal coordinates. With classical statistical mechanics the properties of the system can be derived from this potential energy function. A few new contacts are then introduced as additional energy terms, each having a quadratic dependence on a single internal coordinate. These terms are added as perturbations to the original potential energy, and the attendant changes in the properties of the system are obtained. Exact expressions can be derived for changes in the enthalpy, entropy, and for any arbitrary internal degree of freedom. These quantities are expressed in terms of the parameters of the potential energy functions of the new contacts, and the mean square displacements and positional correlation functions of the internal coordinates. These results provide qualitative insights into the role of contacts in stabilizing a particular conformation. Estimates are given for the entropy of formation of a hydrogen bond in a protein. A criterion is proposed for determining whether a contact is essential to the stability of a protein conformation. This model may be applicable to many experimental systems in which mutant or modified proteins are available that differ by one or a few amino acids. The results may also be useful in thermodynamic analyses of computer simulations.

INTRODUCTION

The molecular interactions that determine the stability of protein conformations and complexes have absorbed the attention of many investigators for many years. Among the theoretical approaches to these questions are attempts to formulate a detailed potential energy and to use a computer to simulate the behavior of a protein (Nemethy and Sheraga, 1977; Levitt, 1982; Karplus and McCammon, 1983). Other theoretical studies have avoided the level of detail used in computer simulations in an attempt to gain some qualitative insights. In one such study Sturtevant (1977) considered many contributions to the heat capacity and entropy changes of a protein, concluding that conformational, solvation (or hydrophobic), and vibrational effects are the most important determinants of protein stability. Cooper and Dryden (1984) proposed a theory for allostery involving ligand-induced changes in vibrational modes. In recent years other theoretical studies (Karplus and Kushick, 1981; Go et al., 1983; Levy et al., 1984) and experimental studies (Hawkes et al., 1984) have arrived at the conclusion that vibrational contributions to the stability of protein conformations are of major importance.

A harmonic potential is a natural theoretical device with which to treat vibrational contributions to the free energy, and has recently been applied to proteins (Levy et al., 1982; Go et al., 1983; Brooks and Karplus, 1983; Levitt et al., 1985). The validity and usefulness of a harmonic potential for proteins have been discussed extensively in the above-cited studies and elsewhere (Bialek and Goldstein, 1985). Although there are limitations, the ease with which a harmonic potential can be used to calculate thermodynamic properties makes it a first order approximation of considerable importance (Karplus and McCammon, 1983).

To gain a better understanding of the contribution of vibrational free energy to the stability of proteins, a theoretical model has been developed with the harmonic potential as its basis. In this model a protein is treated as a collection of atoms connected by perfect harmonic springs. One or a few new springs are added as a perturbation to the system. The entire system responds to this perturbation, and the structural and thermodynamic consequences of these additions are determined. Relations are then derived that do not depend on a detailed knowledge of the form of the potential energy function. Many of these results can be used in conjunction with computer simulations to provide a thermodynamic analysis, in the spirit of some previous studies (Karplus and Kushick, 1981; Levy et al., 1984).

A motivation behind the development of this theory for the perturbation of a harmonic system is the possibility of applications to experiments on modified or mutant forms

BIOPHYS. J. © Biophysical Society · 0006-3495/87/02/313/09 $1.00
Volume 51 February 1987 313-321

of proteins. Studies of denaturation (Matthews et al., 1980; Hecht et al., 1984; Hawkes et al., 1984), structure (Grutter et al., 1979), biological activity (Fersht et al., 1985; Craik et al., 1985; Ackers and Smith, 1985), and subunit dissociation (Ackers and Smith, 1985) of proteins with single or multiple amino acid substitutions have identified some important specific interactions. It is hoped that such substitutions can be modeled as harmonic perturbations, and that the theory presented here may be useful in the interpretation of these kinds of experiments.

THE MODEL

We begin by assuming that the potential energy of a protein is locally quadratic in a set of internal coordinates.

$$V = V_0 + \sum_{i,j}^{n} x_i x_j a_{ij}, \tag{1}$$

where V is the potential energy, x is a set of internal displacement coordinates,[1] and the temperature-independent elements a_{ij} form a symmetric square matrix A. There are N atoms and $n = 3N - 6$ internal degrees of freedom in potential energy; the potential energy at the local minimum is V_0.

There are two strategies for including solvent effects (McCammon, 1984). The model system can include a shell of solvent molecules; the set of internal coordinates would include their positions, and some elements of A would represent solvent–protein and solvent–solvent interaction energies. Alternatively, the potential energy function can be regarded as a potential of mean force that includes solvent effects. (The elements of A would then be temperature dependent.) The theory developed here does not make use of the explicit form of the potential energy of an entire protein; it matters only that Eq. 1 can include solvent effects in a formal sense. Because the development below makes use of a temperature derivative of the partition function, the preferred treatment of the solvent contribution to vibrational effects is the first procedure of including a shell of solvent molecules in the system.

There is a unitary matrix that transforms this system to normal displacement coordinates (Wilson et al., 1955; Califano, 1976). Following this transformation, and switching to matrix and vector notation, the harmonic potential becomes

$$V = V_0 + \psi\Gamma\psi, \tag{2}$$

where ψ is a vector of normal displacement coordinates, and Γ is a diagonal matrix of n positive eigenvalues denoted as γ_i.

The harmonic vibrational contributions to the thermodynamic properties of a system are completely determined by the potential energy as expressed in Eqs. 1 or 2 . If we take the classical limit (see Brooks and Karplus, 1983, and Levy et al., 1984, for a discussion of the classical limit for the vibrational energy of a protein), the partition function, and hence the entropy and enthalpy, can be computed from the classical configuration integral (MacQuarrie, 1976) to give

$$S = -\tfrac{1}{2}nk - \tfrac{1}{2}k \ln|\Gamma| = -\tfrac{1}{2}nk - \tfrac{1}{2}\ln|\mathrm{A}| \tag{3}$$

and

$$H = V_0 + \tfrac{1}{2}nkT, \tag{4}$$

where k is the Boltzmann constant and T is the temperature. In this analysis the number of degrees of freedom, n, does not change; terms with an explicit dependence on the number of degrees of freedom, including the internal kinetic energy, $\tfrac{1}{2}nkT$, will be neglected in subsequent discussions.

The task is now to examine how this system will change if it is modified in such a way that new internal contacts are introduced. It is important to realize that the perturbations introduced here involve only a change in the Hamiltonian and not a change in the number of degrees of freedom. The potential energy, u, of a new internal contact is expressed as a single term with a quadratic dependence on one particular internal coordinate, x_s. Thus,

$$u = u_{s0} + a_s x_s^2, \tag{5}$$

where u_{s0} is the minimum potential energy for this contact, and a_s is twice its force constant; both a_s and u_0 are independent of temperature. x_s is the distance between two interacting atoms minus the distance of minimum potential energy for that particular interaction. In a formal sense, Eq. 5 represents an interaction between two atoms of the structure that was not included in the original harmonic potential in Eq. 1. Generalizing to an arbitrary number of contacts gives

$$u = \sum_s (u_{s0} + a_s x_s^2), \tag{6}$$

where the index s can now be varied to denote different contacts.

The energies of real internal contacts depend on more than one internal degree of freedom. Dependence on an angle as well as a distance is typical. The most important degree of freedom for hydrogen bonds and salt bridges is the bond length. A dependence on angle, or on any other internal degree of freedom, could be approximated by additional terms in Eq. 6. For the sake of simplicity, in this work a single contact will be represented by a single quadratic term.

The addition of the new terms changes the normal coordinates, the matrix Γ, and the position and the potential energy at the minimum. The objective now is to

[1]Throughout this paper displacement means that the coordinate is referred to its position of minimum potential energy.

reevaluate the thermodynamic quantities with Eqs. 3 and 4, after the addition of the new terms represented by Eqs. 5 and 6, to the total potential energy in Eq. 2.

The first step in solving this problem is to express the x_s in terms of the original normal displacement coordinate ψ. Since any internal coordinate can be expressed as a linear combination of the normal coordinates, we have

$$x_s = \psi \cdot \mathbf{g}_s - x_{s0}. \tag{7}$$

x_{s0} gives the position of the minimum potential energy for contact s, relative to the minimum potential energy of the unperturbed structure. The vector $\mathbf{g}_s$ could be a row of the inverse of the matrix used to transform Eq. 1 to Eq. 2. However, sets of internal coordinates are not unique (Wilson et al., 1955) and the internal coordinates used in Eqs. 5 or 6 need not appear in Eq. 1 for Eq. 7 to be valid. Substituting Eq. 7 into Eq. 6 gives

$$u = \psi \mathrm{G} \psi - \sum_s [a_s(2x_{s0}\psi \cdot \mathbf{g}_s - x_{s0}^2) - u_{s0}] \tag{8}$$

where the $n \times n$ matrix G is

$$\mathrm{G} = \begin{pmatrix} \sum_s a_s g_{s1}^2 & \sum_s a_s g_{s2} g_{s1} & \cdots & \sum_s a_s g_{sn} g_{s1} \\ \sum_s a_s g_{s2} g_{s1} & \sum_s a_s g_{s2}^2 & \cdots & \sum_s a_s g_{sn} g_{s2} \\ \vdots & \vdots & & \vdots \\ \sum_s a_s g_{sn} g_{s1} & \sum_s a_s g_{sn} g_{s2} & \cdots & \sum_s a_s g_{sn}^2 \end{pmatrix}. \tag{9}$$

The rank of this matrix is important to subsequent analysis and is examined in Appendix A. The new potential energy is obtained by adding Eq. 8 to Eq. 2 to give

$$V' = V_0 + \psi(\Gamma + \mathrm{G})\psi - \sum_s \{2a_s x_{s0}\mathbf{g}_s \cdot \psi - a_s x_{s0}^2 - u_{s0}\}. \tag{10}$$

The position of the new potential energy minimum, ψ', can be determined by differentiating Eq. 10 with respect to ψ and setting the derivative equal to 0. The resulting relationship is

$$(\Gamma + \mathrm{G})\psi' = \sum_s a_s x_{s0}\mathbf{g}_s \tag{11}$$

ψ' are computed explicitly in Appendix C with Cramer's rule. However, for determining thermodynamic quantities it is more useful to express ψ' as

$$\psi' = (\Gamma + \mathrm{G})^{-1} \sum_s a_s x_{s0}\mathbf{g}_s. \tag{12}$$

Eq. 10 can now be transformed to a new set of displacement coordinates ψ'', referred to the new minimum at ψ'. The new coordinates are not normal coordinates. The transformation gives

$$V' = V_0 + \psi''(\Gamma + \mathrm{G})\psi'' + \sum_s (a_s x_{s0}^2 + u_{s0}) - \left(\sum_s a_s x_{s0}\mathbf{g}_s\right)(\Gamma + \mathrm{G})^{-1}\left(\sum_s a_s x_{s0}\mathbf{g}_s\right). \tag{13}$$

The dependence on position in Eq. 13 is quadratic, making the computation of the classical configuration integral identical to that for Eq. 1 or 2. By analogy with Eqs. 3 and 4 we see that the new entropy and enthalpy of the system, after the incorporation of new internal contacts into its structure, are

$$S' = -\tfrac{1}{2}k \ln|\Gamma + \mathrm{G}| \tag{14}$$

and

$$H' = V_0 - \left(\sum_s a_s x_{s0}\mathbf{g}_s\right)(\Gamma + \mathrm{G})^{-1}\left(\sum_s a_s x_{s0}\mathbf{g}_s\right) + \sum_s (a_s x_{s0}^2 + u_{s0}). \tag{15}$$

From this point the problem is purely mathematical. Appendix B shows how the determinant of the matrix in Eq. 14 can be reduced to an expansion of the form

$$|\Gamma + \mathrm{G}| = |\Gamma|\left\{1 + \sum_s a_s Y_{ss} + \sum_{s>r} a_s a_r \begin{vmatrix} Y_{ss} & Y_{rs} \\ Y_{sr} & Y_{rr} \end{vmatrix} + \sum_{s>r>t} a_s a_r a_t \begin{vmatrix} Y_{ss} & Y_{rs} & Y_{ts} \\ Y_{sr} & Y_{rr} & Y_{tr} \\ Y_{st} & Y_{rt} & Y_{tt} \end{vmatrix} + \cdots\right\}, \tag{16}$$

where the summation index $s > r$ indicates that each pair is only counted once, and where sums of the form $\Sigma_i\, g_{si}\, g_{ri}/\gamma_i$ are represented by the symbol Y_{sr}. These sums reappear in most of the analysis that follows, and are therefore very important. It can be shown that

$$\overline{x_s x_r} = \tfrac{1}{2}kTY_{sr}, \tag{17}$$

where the bar denotes an ensemble average. This is derived by integrating the Boltzmann distribution with the potential taken from Eq. 2. x_s and x_r are expressed in terms of normal coordinates as in Eq. 7, except that since x_s and x_r are from the unperturbed system, $x_{s0} = x_{r0} = 0$. This relation has been used in many other applications. Eq. 17 is especially useful in the quasi-harmonic approximation (Karplus and Kushick, 1981; Levy et al., 1984; Bialek and Goldstein, 1985), and is a special case of more general expressions (Maradudin, 1969; Califano, 1976).

Appendix D explains how the enthalpy can be expressed

as

$$H' = V_0 + \sum_s (a_s x_{s0}^2 + u_{s0}) + \frac{\begin{vmatrix} 0 & \sum_s a_s x_{s0} g_{s1} & \sum_s a_s x_{s0} g_{s2} & \cdots & \sum_s a_s x_{s0} g_{sn} \\ \sum_s a_s x_{s0} g_{s1} & & & & \\ \sum_s a_s x_{s0} g_{s2} & & & & \\ \vdots & & \Gamma + G & & \\ \sum_s a_s x_{s0} g_{sn} & & & & \end{vmatrix}}{|\Gamma + G|} \tag{18}$$

where the denominator in the last term can be taken from Eq. 16 and the numerator can be expressed as an expansion of the form (Appendix D)

$$|\Gamma| \left\{ \sum_s \sum_r a_s a_r x_{s0} \begin{vmatrix} 0 & x_{r0} \\ Y_{sr} & Y_{rr} \end{vmatrix} + \sum_{r>t} \sum_s a_s a_r a_t x_{s0} \begin{vmatrix} 0 & x_{r0} & x_{t0} \\ Y_{rs} & Y_{rr} & Y_{tr} \\ Y_{ts} & Y_{tr} & Y_{tt} \end{vmatrix} + \cdots \right\}, \tag{19}$$

An important property of the matrix G is that its rank is equal to the number of contacts, i.e., the number of terms in Eq. 6 (Appendix A). The number of terms needed in the above expansions is therefore limited by the number of contacts. This point is made in Appendix B for the expansion of $|\Gamma + G|$. Similar arguments apply to the other expansions. This simplifies the results for one contact to

$$S' = S - \tfrac{1}{2} k \ln (1 + a_s Y_{ss}) \tag{20a}$$

and

$$H' = H + a_s x_{s0}^2/(1 + a_s Y_{ss}) + u_{s0}, \tag{20b}$$

where S' and H' are the entropy and enthalpy of the perturbed system. For two contacts, we have

$$S' = S - \tfrac{1}{2} k \ln [(1 + a_s Y_{ss})(1 + a_r Y_{rr}) - a_s a_r Y_{sr}^2] \tag{21a}$$

and

$$H' = H + u_{s0} + u_{r0} + \frac{a_s x_{s0}^2 (1 + a_r Y_{rr}) + a_r x_{r0}^2 (1 + a_s Y_{ss}) - 2 x_{s0} x_{r0} a_s a_r Y_{rs}}{(1 + a_s Y_{ss})(1 + a_r Y_{rr}) - a_s a_r Y_{sr}^2}, \tag{21b}$$

where the subscripts s and r denote parameters for the two different contacts. For three or more contacts, expressions can also be written down. They are complicated, but they are explicit functions of the same quantities that appear in Eqs. 21. In general, the entropy depends only on a_s, Y_{ss}, Y_{sr}, and analogous quantities for each new contact and pair of contacts. The enthalpy depends on these same quantities and, in addition, on the quantities x_{s0}, u_{s0}, as well as analogous quantities for other contacts.

Explicit expressions for changes in any internal coordinate can also be obtained. Another internal coordinate, x_t, which is not involved in any new contact formation, can still be expressed in terms of the normal displacement coordinates with Eq. 7. The change in x_t can be obtained by substituting the position of the new potential energy minimum (ψ' from Appendix C) into an equation analogous to Eq. 7 to give for a single contact

$$x_t' = a_s x_{s0} Y_{ts}/(1 + a_s Y_{ss}) \tag{22a}$$

For two contacts, denoted by the subscripts r and s, we have

$$x_t' = \frac{Y_{rt}[a_r x_{r0} + a_s a_r (x_{r0} Y_{rr} - x_{s0} Y_{sr})] + Y_{st}[a_s x_{s0} + a_s a_r (x_{s0} Y_{ss} - x_{r0} Y_{sr})]}{(1 + a_s Y_{ss})(1 + a_r Y_{rr}) - a_s a_r Y_{sr}^2} \tag{22b}$$

For three or more contacts the expressions are complex. However, as with the thermodynamic quantities, they depend on the same quantities, a_s, x_{s0}, Y_{ss}, Y_{sr}, and analogous quantities for other contacts.

The change in x_s, the coordinate participating in a new contact, is easily obtained from Eqs. 22 by setting all subscripts t equal to s. This gives

$$x_s' = a_s x_{s0} Y_{ss}/(1 + a_s Y_{ss}) \tag{23a}$$

and

$$x_s' = \frac{a_r x_{r0} Y_{rs} + a_s x_{s0} Y_{ss} + a_s a_r x_{s0} (Y_{rr} Y_{ss} - Y_{sr}^2)}{(1 + a_s Y_{ss})(1 + a_r Y_{rr}) - a_s a_r Y_{sr}^2} \tag{23b}$$

for one and two contacts, respectively.

APPLICATIONS

The results of the analysis presented here may be useful in identifying and evaluating the harmonic contribution to the vibrational free energy of a protein. A major factor upon which the application of this theory to mutant or modified proteins depends is how effectively can a change in the potential energy surface of a modified or perturbed protein be represented by quadratic terms of the form in Eqs. 5 and 6. Hydrogen bonds or salt bridges should be well represented by one, or possibly two terms. Steric repulsions and hydrophobic interactions may require several terms.

The ideal substitution is exemplified by interchanging a serine and a cysteine (that does not form a disulfide bond). In this case the energy change could be expressed in terms of a single internal degree of freedom.

In real protein modifications there is often a change in the number of atoms, and thus in the number of internal degrees of freedom. The model developed here is applicable to a system modified in such a way that the number of degrees of freedom does not change. This problem could be avoided by the use of residue or sidechain potentials (Nemethy and Sheraga, 1977), or by the approximation of the potential energies of two different sidechains in terms of the positions of an equivalent number of extended atoms.

The model and relationships developed here may be applicable to some of the many experiments that can be done on mutant or modified proteins. The quantities that appear in these relationships are enthalpy, entropy, internal coordinates, the mean square displacements of internal coordinates, and the parameters of the potential energy functions of new internal contacts. The correlation functions for the displacements of pairs of internal coordinates also appear in this theory. These quantities cannot be measured at present, but can be estimated from computer simulations (Levitt et al., 1985).

Enthalpy and entropy differences between a folded and denatured state can be measured (Matthews et al., 1980; Hawkes et al., 1984; Hecht et al., 1984). The model presented here cannot be applied to the denatured state of a protein. Application of this theory to denaturation experiments may be possible if the effect of modification on the entropy and enthalpy of the denatured state can be estimated from solvation energies (Nozaki and Tanford, 1971; Eisenberg and McLachlan, 1986).

The parameters for the internal contact energy functions that appear in Eqs. 5 or 6 can be taken from potentials used in computer studies (Nemethy and Sheraga, 1977; McCammon et al., 1979; Levitt, 1982; Karplus and McCammon, 1983). Changes in an internal coordinate can be taken from x-ray crystallography. Spectroscopic techniques, especially NMR (Wuthrich, 1982), can also be used to provide this information. Mean square deviations in position can be estimated from Debye-Waller factors (Debrunner and Frauenfelder, 1982; Petsko and Ringe, 1984), although mean square deviations in the positions of atoms are not simply related to the mean square displacement of the internal coordinate that appears in Y_{ss}. Spectroscopic techniques may also be useful in providing mean square displacements (Debrunner and Frauenfelder, 1982; McCammon, 1984).

As an illustration of a simple qualitative application we can ask what range of entropy changes might be expected for the formation of a typical hydrogen bond in a protein. It is assumed that the hydrogen bond is buried within the protein so that the groups involved cannot form hydrogen bonds with water. Eq. 20a can be used if the force constant for a hydrogen bond is available, and if the mean square displacement of the distance between the two interacting atoms is known in the absence of the hydrogen bond. The "10–12 potential" is often used to represent the potential energy as a function of length for a hydrogen bond (Nemethy and Sheraga, 1977; McCammon et al., 1979; Levitt, 1982; Karplus and McCammon, 1983). For a potential of this form the force constant $a = -60\ E_{min}/r_{min}^2$. Values used previously for E_{min} and r_{min} are 3.5 kcal/mol and 2.8 Å, respectively (McCammon et al., 1979). This gives $a = 27$ kcal/Å^2 mol. Mean square displacements of atomic positions in a protein interior range from 0.05 to 0.15 Å^2 (Debrunner and Frauenfelder, 1982), with larger fluctuations occurring near the protein surface or near a terminus. It will be assumed that the positions of the two oxygen atoms involved in the hydrogen bond are uncorrelated in the absence of the hydrogen bond. This leads to an increase by a factor of $\sqrt{2}$. Using these numbers gives a range of from -2.4 to -3.5 entropy units for the formation of a hydrogen bond at 300°K. The entropy change would be substantially smaller if the relative positions of the two oxygen atoms were already constrained by contacts between nearby groups.

An interesting insight into protein stability can be gained by taking a slight twist to the model as formulated and asking what is the consequence of removing a contact from a structure, i.e., instead of adding a spring to the structure, a spring is removed. This amounts to subtracting terms of the form in Eqs. 5 and 6 from the original total potential instead of adding them. The resulting entropy change for breaking a single contact is then $-\frac{1}{2}k\ln(1 - aY)$. Since Y is always positive, the entropy change is always positive, but there is a divergence at $aY = 1$. When aY is >1, the matrix that results from subtracting G from Γ has one or more negative eigenvalues. Within the context of this model, this divergence means that without the contact, there is no longer a local potential energy minimum near the original configuration. By removal of the local minimum the harmonic potential might be invalidated. It is possible that the system could then undergo a change to a new conformation.

Without knowing the fate of the system once the minimum is removed, consideration of this divergence offers a means of evaluating the essentiality of an intramolecular contact to the structure of a macromolecule. By comparing the mean square displacement of the distance between the two interacting atoms with $kT/2a$, the importance of a contact to the conformation of a protein can be assessed. By this criterion, at 300°K a hydrogen bond with a force constant a of 27 kcal/Å^2 mol is judged to be essential if the mean square displacement in bond length is >0.012 Å^2. Unfortunately, measurements of mean square bond length displacements of this magnitude are not experimentally possible. The criterion proposed here for the essentiality of

an internal contact could be applied to proteins with computer simulation techniques.

DISCUSSION

Even without detailed knowledge of the quantities that appear in Eqs. 20 and 21, some useful qualitative generalizations can be made. Because Y_{ss} is always positive, Eq. 20a indicates that the entropy is always reduced after the introduction of a new contact. This is in keeping with the general result that introducing a new constraint into a harmonic system always increases vibrational frequencies (Wilson et al., 1955).

It is instructive to ask what are the limits to the magnitude of the entropy decrease after the formation of a single contact. The sum $\Sigma_i g_{si}^2$ is equal to one (the matrix for the transformation from A to Γ is unitary), so an upper bound to Y_{ss} is the reciprocal of the smallest element of Γ, which will be denoted as ζ. Thus, according to Eq. 20a the entropy can decrease by as much as $-\frac{1}{2}k \ln (1 + a/\zeta)$. The situation represented by this extreme is one where the internal degree of freedom of the new contact overlaps with a very weakly constrained internal mode of the protein: the force constant of the new contact then simply adds to the force constant of that weak normal mode. The largest changes in entropy would then occur when the newly formed contact constrains the softest internal mode of the protein. If the force constant of the new contact, a, is 1,000 times larger than the force constant of the soft mode that is constrained by the new contact, the entropy change will be -6.9 entropy units, and the free energy increase at 300°K will be 2 kcal/mol. If the modes with which the contact interacts are of similar strength, one might expect aY_{ss} to be of the order of 1. In this case the free energy increase would be ~0.2 kcal/mol.

Single amino acid replacements change the entropy of denaturation of some mutant forms of lysozyme by as much as 30 entropy units (Hawkes et al., 1984). This is much larger than can be accounted for by solvation differences between the single amino acid replacements (Nozaki and Tanford, 1971; Eisenberg and McLachlan, 1986). Representing the changes produced by such modifications within the theoretical framework developed here would require several terms in Eq. 6. Otherwise, values of a single force constant or a prior mean square displacement would have to be unrealistically high. It is therefore likely that the single amino acid replacements that produce such large entropy changes involve perturbations of several internal degrees of freedom.

A consequence of Eq. 20b is that the enthalpy change after the addition of a single new contact is bounded below by u_{s0} and above by $u_{s0} + a_s x_{s0}^2$. (It should be noted that x_{s0} is the position of the energy minimum of the new contact relative to the unperturbed structure. The actual value of x_s' in the new structure is between 0 and x_{s0}.) The lower bound of the enthalpy is realized when the new contact fits into the original structure perfectly, so that $x_s' = 0$. The upper bound is realized if the protein retains its original structure, and the only strain appears in the new contact, so that $x_s' = x_{s0}$. Examination of Eqs. 20a and 20b shows that the actual value of the enthalpy within these bounds is completely determined by x_s' and the entropy change.

It can be shown that all of the terms of the expansion of $|\Gamma + G|$ in Eq. 16 are always positive (Hardy et al., 1952), and therefore have a lower bound of zero. This means that the upper bound to a change in entropy after the addition of several new contacts is $-\frac{1}{2}k \ln (1 + \Sigma_s a_s Y_{ss})$. A lower bound to the entropy change after the addition of several contacts is derived by replacing the terms of the expansion in Eq. 16 by their upper bounds. The upper bound of the determinant of a matrix of the form that appears in this expansion is the product of the diagonal elements (Hardy et al., 1952). After this replacement, $|\Gamma + G|/|\Gamma|$ can be factored into $\Pi_s (1 + a_s Y_{ss})$. Thus, the lower bound to the change in entropy is the sum of all the changes after the addition of each contact separately.

The limits for the entropy changes can be summarized as follows. The maximal decrease in entropy is realized when there is no overlap between the various degrees of freedom affected by the various new contacts. The minimal decrease in entropy occurs when there is complete redundancy. Thus, entropic destabilization of a structure is best achieved by distributing new contacts over as many normal modes as possible.

The degree of overlap between different internal degrees of freedom is reflected in the magnitude of Y_{sr}. When Y_{sr} is zero, Eqs. 21 can be used to show that the changes in entropy and enthalpy produced by incorporating two new contacts decompose into the sum of the changes resulting from incorporating each individual contact. Ackers and Smith (1985) reviewed studies in which proteins with single and double modifications are compared. The majority of the cases considered show additive effects of two modifications. In a few instances double modifications were not additive, suggesting that there is some overlap between the degrees of freedom perturbed by the two modifications. To the extent that a single modification can be represented by the addition of one quadratic term, the results of such experiments could be used to estimate positional correlations. Future efforts will be directed towards improving the theory for cases where each modification is represented by several quadratic terms (Jackson, M. B., manuscript submitted for publication).

It is notable that only pairwise positional correlation functions (Eq. 17) contribute to the above expressions, and not higher order correlation functions. This follows from the assumption of a quadratic form for the potential energy in Eqs. 1, 2, and 6. An important consequence of this is that if one has a set of contacts to examine, once all single and pairwise experiments have been carried out, results for any

higher order combination of the same perturbations can be predicted. The single and pairwise terms completely define the system.

APPENDIX A

Rank of the Matrix G in Eq. 9

If only a single contact is included, then any 2 × 2 submatrix of G will be of the form

$$\begin{pmatrix} a_s g_{si}^2 & a_s g_{si} g_{sj} \\ a_s g_{sj} g_{si} & a_s g_{sj}^2 \end{pmatrix}. \tag{A1}$$

Factoring out g_{si} from the first column and g_{sj} from the second column leaves a matrix with two identical columns, hence its determinant is zero. Thus, with only a single contact, the rank of G is one.

For an arbitrary number of contacts, j, we can represent a $k \times k$ submatrix of G by its columns

$$\mathrm{G} = \left(\sum_{s_1} a_{s_1} g_{s_1 1} \mathbf{g}_{s_1}, \sum_{s_2} a_{s_2} g_{s_2 2} \mathbf{g}_{s_2}, \ldots, \sum_{s_k} a_{s_k} g_{s_k k} \mathbf{g}_{s_k} \right), \tag{A2}$$

where s_i is an index that is varied over all terms of Eq. 6. It is undestood that each column, $\mathbf{g}_s$, has been truncated to k elements. The determinant can then be written as the multiple sum

$$\sum_{s_1} \sum_{s_2} \sum_{s_3} \cdots \sum_{s_k} \left| a_{s_1} g_{s_1 1} \mathbf{g}_{s_1}, a_{s_2} g_{s_2 2} \mathbf{g}_{s_2}, \ldots, a_{s_k} g_{s_k k} \mathbf{g}_{s_k} \right|. \tag{A3}$$

If k, the dimension of the submatrix, is greater than j, the number of contacts, then every term in Eq. A3 will be zero because in at least two columns, the value of s_i will be the same and the columns will be linearly dependent. Only when each column is derived from a different contact will all the s_i be different, producing a nonzero term. Thus, the highest order nonsingular submatrix of G is of order j.

APPENDIX B

Determinant of Γ + G

The determinant of a matrix that is the sum of a diagonal matrix and a nondiagonal matrix can be expanded as (Aitken, 1956)

$$|\Gamma + \mathrm{G}| = |\Gamma| \left\{ 1 + \sum_i \frac{\mathrm{G}_{ii}}{\gamma_i} + \sum_{i>j} \frac{\begin{vmatrix} \mathrm{G}_{ii} & \mathrm{G}_{ij} \\ \mathrm{G}_{ji} & \mathrm{G}_{jj} \end{vmatrix}}{\gamma_i \gamma_j} + \sum_{i>j>k} \frac{\begin{vmatrix} \mathrm{G}_{ii} & \mathrm{G}_{ij} & \mathrm{G}_{ik} \\ \mathrm{G}_{ji} & \mathrm{G}_{jj} & \mathrm{G}_{jk} \\ \mathrm{G}_{ki} & \mathrm{G}_{kj} & \mathrm{G}_{kk} \end{vmatrix}}{\gamma_i \gamma_j \gamma_k} + \cdots \right\}, \tag{B1}$$

where $\mathrm{G}_{ij} = \Sigma_s a_s g_{si} g_{sj}$. Note that each term in the expansion has successively higher order submatrices of G. Appendix A shows how the submatrices of G of order greater than the number of contacts have determinants of 0. Thus the number of terms needed in the expansion is limited by the number of contacts in the perturbation.

The first order term in Eq. B1 is simply $\Sigma_s a_s Y_{ss}$. The second order term is

$$\sum_{rs} \sum_{i>j} \frac{\begin{vmatrix} a_s g_{si}^2 & a_r g_{ri} g_{rj} \\ a_s g_{si} g_{sj} & a_r g_{rj}^2 \end{vmatrix}}{\gamma_i \gamma_j}, \tag{B2}$$

where s and r are two independently varied indices that denote the contact. Factoring out $a_s g_{si}$ from column one and $a_r g_{rj}$ from column two, and multiplying row one by g_{si}/γ_i and row two by g_{rj}/γ_j gives

$$\sum_{s>r} a_s a_r \begin{vmatrix} \sum_i g_{si}^2/\gamma_i & \sum_i g_{si} g_{ri}/\gamma_i \\ \sum_j g_{rj} g_{sj}/\gamma_j & \sum_j g_{rj}^2/\gamma_j \end{vmatrix} = \sum_{s>r} a_s a_r \begin{vmatrix} Y_{ss} & Y_{rs} \\ Y_{sr} & Y_{rr} \end{vmatrix}. \tag{B3}$$

Similar manipulations allow the third order term to be expressed as

$$\sum_{s>r>t} a_s a_r a_t \begin{vmatrix} Y_{ss} & Y_{rs} & Y_{ts} \\ Y_{sr} & Y_{rr} & Y_{tr} \\ Y_{st} & Y_{rt} & Y_{tt} \end{vmatrix}. \tag{B4}$$

The general form of the expansion is evident. Eqs. B3 and B4 can also be obtained by a more complex derivation that makes use of the Cauchy-Binet theorem (Aitken, 1956).

APPENDIX C

Position of the New Potential Energy Minimum

Applying Cramer's rule to Eq. 11 gives

$$\psi_i' = \frac{\begin{vmatrix} \gamma_1 + \mathrm{G}_{11} & \mathrm{G}_{21} & \cdots & \sum_s a_s x_{s0} g_{1s} & \cdots & \mathrm{G}_{n1} \\ \mathrm{G}_{11} & \gamma_2 + \mathrm{G}_{22} & \cdots & \sum_s a_s x_{s0} g_{2s} & \cdots & \mathrm{G}_{n2} \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \mathrm{G}_{1i} & \mathrm{G}_{2i} & \cdots & \sum_s a_s x_{s0} g_{is} & \cdots & \mathrm{G}_{ni} \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \mathrm{G}_{1n} & \mathrm{G}_{2n} & \cdots & \sum_s a_s x_{s0} g_{in} & \cdots & \gamma_n + \mathrm{G}_{nn} \end{vmatrix}}{|\Gamma + \mathrm{G}|} \tag{C1}$$

Applying the expansion of the determinant of a sum of a diagonal and a nondiagonal matrix (Aitken, 1956) gives

$$\frac{|\Gamma|}{|\Gamma+\mathrm{G}|}\left\{\sum_s a_s x_{s0} g_{is}/\gamma_i + \sum_j \frac{\begin{vmatrix} \mathrm{G}_{jj} & \sum_s a_s x_{s0} g_{js} \\ \mathrm{G}_{ij} & \sum_s a_s x_{s0} g_{js} \end{vmatrix}}{\gamma_i\gamma_j} + \sum_{j>k} \frac{\begin{vmatrix} \mathrm{G}_{jj} & \mathrm{G}_{kj} & \sum_s a_s x_{s0} g_{js} \\ \mathrm{G}_{jk} & \mathrm{G}_{kk} & \sum_s a_s x_{s0} g_{ks} \\ \mathrm{G}_{ji} & \mathrm{G}_{ki} & \sum_s a_s x_{s0} g_{is} \end{vmatrix}}{\gamma_i\gamma_j\gamma_k} + \cdots \right\}. \quad \text{(C2)}$$

Factoring out by columns, multiplying the appropriate rows by the appropriate factors, and resumming in a fashion similar to that used in Appendix B, we have

$$\sum_r \sum_s \frac{a_s a_r x_{s0}}{\gamma_i} \begin{vmatrix} g_{si} & g_{ri} \\ Y_{sr} & Y_{ss} \end{vmatrix}$$

$$\text{and} \sum_{r>t} \sum_s \frac{a_s a_r a_t x_{s0}}{\gamma_i} \begin{vmatrix} g_{si} & g_{ri} & g_{ti} \\ Y_{sr} & Y_{rr} & Y_{tr} \\ Y_{st} & Y_{rt} & Y_{tt} \end{vmatrix} \quad \text{(C3)}$$

as the second and third terms, respectively.

To determine a specific internal coordinate, x_q, Eq. 7 is used. The first term is simply

$$\sum_s a_s Y_{sq}. \quad \text{(C4)}$$

The second and third terms are

$$\sum_r \sum_s a_s a_r x_{s0} \begin{vmatrix} Y_{sq} & Y_{rq} \\ Y_{sr} & Y_{ss} \end{vmatrix}$$

$$\text{and} \sum_{r>t} \sum_s a_s a_r a_t x_{s0} \begin{vmatrix} Y_{sq} & Y_{rq} & Y_{tq} \\ Y_{sr} & Y_{rr} & Y_{tr} \\ Y_{st} & Y_{rt} & Y_{tt} \end{vmatrix}, \quad \text{(C5)}$$

and the general form of the expansion is evident.

APPENDIX D

Calculation of the Enthalpy

It will first be shown that the second term on the right of Eq. 15 and the third term on the right of Eq. 18 are identical. Partitioning the numerator of the third term of Eq. 18 as

$$\begin{vmatrix} 0 & \boldsymbol{x} \\ \boldsymbol{x} & \Gamma+\mathrm{G} \end{vmatrix} = \begin{vmatrix} 1 & \boldsymbol{x} \\ \boldsymbol{x} & \Gamma+\mathrm{G} \end{vmatrix} - |\Gamma+\mathrm{G}|.$$

Then using

$$\begin{vmatrix} 1 & -\boldsymbol{x}(\Gamma+\mathrm{G})^{-1} \\ 0 & I \end{vmatrix} \begin{vmatrix} 1 & \boldsymbol{x} \\ \boldsymbol{x} & \Gamma+\mathrm{G} \end{vmatrix} \begin{vmatrix} 1 & 0 \\ -(\Gamma+\mathrm{G})^{-1}\boldsymbol{x} & I \end{vmatrix} = \begin{vmatrix} 1-\boldsymbol{x}(\Gamma+\mathrm{G})^{-1}\boldsymbol{x} & 0 \\ 0 & \Gamma+\mathrm{G} \end{vmatrix}$$

gives

$$\frac{\begin{vmatrix} 0 & \boldsymbol{x} \\ \boldsymbol{x} & \Gamma+\mathrm{G} \end{vmatrix}}{|\Gamma+\mathrm{G}|} = \frac{\begin{vmatrix} 1-\boldsymbol{x}(\Gamma+\mathrm{G})^{-1}\boldsymbol{x} & 0 \\ 0 & \Gamma+\mathrm{G} \end{vmatrix} - |\Gamma+\mathrm{G}|}{|\Gamma+\mathrm{G}|}$$

$$= -\boldsymbol{x}(\Gamma+\mathrm{G})^{-1}\boldsymbol{x}.$$

This can also be seen by using the Cauchy expansion (Aitken, 1956) of the determinant in the numerator of the last term in Eq. 18.

A diagonal expansion of the numerator of Eq. 18 gives

$$\sum_i \begin{vmatrix} 0 & \sum_s a_s x_{s0} g_{si} \\ \sum_s a_s x_{s0} g_{si} & \mathrm{G}_{ii} \end{vmatrix} + \sum_{i>i} \begin{vmatrix} 0 & \sum_a a_s x_{s0} g_{si} & \sum_s a_s x_{s0} g_{sj} \\ \sum_s a_s x_{s0} g_{si} & \mathrm{G}_{ii} & \mathrm{G}_{ji} \\ \sum_s a_s x_{s0} g_{sj} & \mathrm{G}_{ij} & \mathrm{G}_{jj} \end{vmatrix}. \quad \text{(D1)}$$

Manipulations similar to those used in Appendices B and C reduce this expression to

$$\sum_{sr} a_s a_r x_{s0} \begin{vmatrix} 0 & x_{r0} \\ Y_{sr} & Y_{rr} \end{vmatrix} + \sum_{r>t} \sum_s a_s a_r a_t x_{s0} \begin{vmatrix} 0 & x_{r0} & x_{t0} \\ Y_{rs} & Y_{rr} & Y_{rt} \\ Y_{st} & Y_{tr} & Y_{tt} \end{vmatrix}. \quad \text{(D2)}$$

I thank Bill Gelbart, David Eisenberg, and Doug Rees for helpful discussions and critical readings of this manuscript.

Received for publication 8 April 1986 and in final form 15 September 1986.

REFERENCES

Ackers, G. K., and F. R. Smith. 1985. Effects of site specific amino acid modification on protein interactions and biological function. *Annu. Rev. Biochem.* 54:597–629.

Aitken, A. C. 1956. Determinants and Matrices, Greenwood Press, Westport, CT.

Bialek, W., and R. F. Goldstein. 1985. Do vibrational spectroscopies uniquely describe protein dynamics? The case for myoglobin. *Biophys. J.* 48:1027–1044.

Brooks, B., and M. Karplus. 1983. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA.* 80:6571–6575.

Califano, S. 1976. Vibrational States. John Wiley & Sons, New York.
Cooper, A., and D. T. F. Dryden. 1984. Allostery without conformational change. A plausible model. *Eur. J. Biophys.* 11:103–110.
Craik, C. S., C. Largman, T. Fletcher, S. Roczniak, P. J. Barr, R. Fletterick, and W. J. Rutter. 1985. Redesigning trypsin: alteration of specificity. *Science* (*Wash. DC*). 228:291–297.
Debrunner, P. G., and H. Frauenfelder. 1982. Dynamics of proteins. *Annu. Rev. Phys. Chem.* 33:283–299.
Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature* (*Lond.*). 319:199–203.
Fersht, A. R., J.P. Shi, J. Knill-Jones, D. M. Lowe, A. J. Wilkinson, D. M. Blow, P. Brick, P. Carter, M. M. Y. Waye, and G. Winter. 1985. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* (*Lond.*). 314:235–238.
Go, N., T. Noguti, and T. Nishikawa. 1983. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA.* 80:3696–3700.
Grutter, M. G., R. B. Hawkes, and B. W. Hawkes. 1979. Molecular basis of thermostability in the lysozyme from bacteriophage T4. *Nature* (*Lond.*). 277:667–669.
Hardy, G. H., J. E. Littlewood, and G. Polya. 1952. Inequalities. Cambridge University Press, Cambridge. 16, 34.
Hawkes, R., M. G. Grutter, and J. Schellman. 1984. Thermodynamic stability and point mutations of bacteriophage T4 lysozyme. *J. Mol. Biol.* 175:195–212.
Hecht, M. H., J. M. Sturtevant, and R. T. Sauer. 1984. Effects of single amino acid replacements on the thermal stability of the NH_2 terminal domain of phage λ repressor. *Proc. Natl. Acad. Sci. USA.* 81:5685–5689.
Karplus, M., and J. N. Kushick. 1981. Methods for estimating the configurational entropy of macromolecules. *Macromolecules.* 14:325–332.
Karplus, M., and J. A. McCammon. 1983. Dynamics of proteins: elements and function. *Annu. Rev. Biochem.* 52:263–300.
Levitt, M. 1982. Protein conformation, dynamics, and folding by computer simulation. *Annu. Rev. Biophys. Bioeng.* 11:251–271.
Levitt, M., C. Sander, and P. S. Stern. 1985. Protein normal mode dynamics: trypsin inhibitor, crambin, ribonuclease, and lysozyme. *J. Mol. Biol.* 181:423–447.
Levy, R. M., M. Karplus, J. Kushick, and D. Perahia. 1984. Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an α-helix. *Macromolecules.* 17:1370–1374.
Levy, R. M., D. Perahia, and M. Karplus. 1982. Molecular dynamics of an α-helical polypeptide: temperature dependence and deviation from harmonic behavior. *Proc. Natl. Acad. Sci. USA.* 79:1346–1350.
Maradudin, A. A. 1969. Lattice Dynamics. The Benjamin Co. Inc., Elmsford, NY.
Matthews, C. R., M. M. Crisanti, G. L. Gepner, G. Velicelibi, and J. M. Sturtevant. 1980. Effect of single amino acid substitutions on the thermal stability of the α-subunit of tryptophan synthetase. *Biochemistry.* 19:1290–1293.
McCammon, J. A. 1984. Protein dynamics. *Rep. Prog. Phys.* 47:1–46.
McCammon, J. A., P. G. Wolynes, and M. Karplus. 1979. Picosecond dynamics of tyrosine side chains in proteins. *Biochemistry.* 18:927–942.
MacQuarrie, D. 1976. Statistical Mechanics. Harper & Row Publishers Inc., New York.
Nemethy, G., and H. Sheraga. 1977. Protein folding. *Q. Rev. Biophys.* 10:239–352.
Noguti, T., and N. Go. 1982. Collective variable description of small amplitude conformational fluctuations in globular proteins. *Nature* (*Lond.*). 296:776–778.
Nozaki, Y., and C. Tanford. 1971. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. *J. Biol. Chem.* 246:2211–2217.
Petsko, G. A., and D. Ringe. 1984. Fluctuations in protein structure from x-ray diffraction. *Annu. Rev. Biophys. Bioeng.* 13:331–371.
Sturtevant, J. M. 1977. Heat capacity and entropy changes in processes involving proteins. *Proc. Natl. Acad. Sci. USA.* 74:2236–2240.
Wilson, E. B., J. C. Decius, and P. C. Cross. 1955. Molecular Vibrations. McGraw-Hill Book Co., New York.
Wurthrich, K. 1982. High resolution NMR experiments for studies of protein conformations. *In* Structural Molecular Biology Methods and Applications. D. B. Davies, W. Saenger, and S. S. Danyluk, editors. Plenum Publishing Corp., New York. 215–235.